OPINION Epigenome dynamics

Johannes, Frank; Colot, Vincent; Jansen, Ritsert C.

Published in:
Nature Reviews Genetics

DOI:
10.1038/nrg2467

Publication date:
2008

OPINION

Epigenome dynamics: a quantitative genetics perspective

Frank Johannes, Vincent Colot and Ritsert C. Jansen

Abstract | Classically, quantitative geneticists have envisioned DNA sequence variants as the only source of heritable phenotypes. This view should be revised in light of accumulating evidence for widespread epigenetic variation in natural and experimental populations. Here we argue that it is timely to consider novel experimental strategies and analysis models to capture the potentially dynamic interplay between chromatin and DNA sequence factors in complex traits.

The heritable basis of complex traits has long been assumed to rest on the stable transmission of multiple causative DNA sequence alleles from parents to offspring. This classical view is being challenged by the recent discovery that variation in chromatin states is highly abundant in experimental and natural populations[1–5], and could therefore provide an additional source of phenotypic variation. Indeed, chromatin differences between individuals can exist independently of DNA sequence polymorphisms and can be transmitted across mitosis and meiosis in both mammals and plants, with phenotypic consequences at the level of the cell, tissue or whole organism[6–10].

When alternative chromatin states (epialleles) are stable across generations, they are functionally indistinguishable from DNA sequence alleles at the population level. Several examples of such stably transmitted epialleles have been discovered in plants[11–13], but were originally thought to be DNA variants. More commonly, however, meiotically transmitted epialleles display intriguing patterns of instability[14], which can be induced by genomic and environmental perturbations[7]. Newly acquired epiallelic forms can remain stable or revert over one or several generations in a manner that is either dependent or independent of the nucleotide sequence[7,15,16] (F.J. and V.C., unpublished observations). These dynamic features are not easily integrated into our current understanding of how complex traits are created and sustained in populations. Clearly, for chromatin variation to be formally incorporated into our existing quantitative genetic framework, it is necessary to obtain a genome-wide characterization of its temporal properties. In particular, we require a clear picture of the transgenerational relationship between DNA sequence alleles and epialleles, and estimates of the relative contributions of these two sources of variation to heritable phenotypes[17].

Here we argue that this goal can be achieved within a QTL mapping framework using multigenerational data derived from natural or experimental populations of genetically diverse individuals. By treating epialleles as generation-dependent molecular phenotypes, we show how to map genome-wide DNA sequence variants that can modulate their dynamic behaviour across generations in *cis* or in *trans*. This approach can be used to uncover the autonomous (epigenetic) aspect of the chromatin inheritance system, which requires decomposition of the epigenome into sequence-dependent and sequence-independent regions of variation and stability. The resulting information will provide a means to begin delineating phenotypic variation into several components: a component originating from DNA sequence variation, a component originating from epigenetic variation, a component of non-heritable chromatin variation and a component of unexplained (residual) variation.

We illustrate through simple examples how this approach can be used not only in transgenerational contexts (meiotic experiments) but also in intragenerational contexts (mitotic experiments), and we stress that the inclusion of both chromatin and DNA sequence data may be necessary to dissect the potentially dynamic architecture of complex traits.

Experimental strategies

Single gene perturbations. With the advent of high-resolution, genome-wide measurement technologies, such as chromatin immunoprecipitation coupled with hybridization to tiling arrays (ChIP–chip) or with deep sequencing (ChIP–seq), it has become feasible to construct genome-wide chromatin maps for various organisms[18–20]. These technologies have recently been used to explore the epigenomic landscapes in several species[15,21–26]. The most commonly assayed chromatin marks are DNA methylation and various post-translational modifications of histone proteins. Attempts at relating chromatin variation to DNA sequence variation in experimental populations are currently limited to forward genetic strategies. In these settings, the function of a single gene or a small set of genes is disrupted (that is, made variable) and the consequences on the epigenome are tracked. Recent comparisons, for example, of the genome-wide DNA methylation profiles of wild-type *Arabidopsis thaliana* plants with that of mutants lacking genes important for DNA methylation have revealed markedly altered methylation states at several hundred genes located in *trans* to the conditioning loci[15,21,24,27]. The extension of this approach to a global (unbiased) search for genetic loci that can affect DNA methylation or other chromatin modifications has known limitations[28]: it would require the successive perturbation of all genes and combinations thereof, followed by a global epigenomic assessment of each perturbation — a task that is practically and conceptually infeasible[29].

Multigenic perturbations. A powerful alternative is the use of experimental populations derived from crosses of different inbred parents. Recently, Zhang and

colleagues[2] followed the genome-wide DNA methylation profiles of two *A. thaliana* natural accessions to their filial 1 (F_1) offspring and showed that the transgenerational inheritance of DNA methylation states occurred in a predominantly additive manner, with the F_1 individuals displaying values intermediate between the two parents at each sampled genomic location. This result is consistent with previous reports[1,3,4], and suggests that the transmission of epialleles can be remarkably stable in plant populations. However, to distinguish whether this stability is conferred by sequence-independent chromatin inheritance (epigenetics) or by *cis*- or *trans*-acting DNA-based factors requires the construction of more advanced crosses (such as an F_2 generation or recombinant inbred lines (RILs)) coupled with QTL mapping methods.

Riddle and Richards[3] exploited this fact, albeit on a more limited scale, in their analysis of RILs derived from two *A. thaliana* parents that differed substantially in their level of DNA methylation in nucleolus organizer regions (NORs) located at the tip of chromosome 2 and of chromosome 4. By treating interindividual differences in DNA methylation in NORs as a molecular quantitative trait, they showed that 50% of the variation in the population was related to QTLs linked to markers that mapped in *cis* to the NOR, whereas 20% was explained by *trans*-acting loci. Interestingly, the authors showed that part of the *cis* effect was probably attributable to the inheritance of parental DNA methylation profiles. This study demonstrated, in principle, that regional chromatin variation can indeed be resolved into autonomous as well as *cis* or *trans* DNA sequence-dependent components.

A proposed global and transgenerational approach. We argue that the scope of this type of decomposition should be broadened to include system-wide measurements on the same individuals in the population and their offspring. This involves genome-wide profiling of DNA sequence and chromatin variation as well as higher-level phenotypic information. By formally tracing the relationship of these three levels of analysis through genetically informative pedigrees (that is, multiple generations), it is possible to simultaneously delineate the relationship between DNA sequence alleles and epialleles in a locus-by-locus manner (both *cis* and *trans*), and to estimate their respective effects on transmitted phenotypes. This can be achieved with experimental populations (such as RILs or F_2 offspring) but also with natural ones (for example, humans), as long as pedigrees can be sampled to allow for the observation of different inheritance patterns. An appealing experimental starting point, and the focus of this discussion, is existing populations of RILs (FIG. 1a). Because DNA sequence remains virtually stable following further propagation of an individual line, subtle chromatin dynamics across generations in a meiotic experiment can be systematically assessed against the fixed genotypic background of that line, and across a range of different DNA backgrounds at the population level (FIG. 1b).

It is important to note that the proposed approach can be also applied to developmental contexts in mitotic experiments, as it could be relevant in cell differentiation and cancer studies (FIG. 1b). In this case, the time-dependent nature of chromatin stability needs to be framed in terms of mitotic rather than meiotic transitions.

In the following sections we explore the implications of this new experimental vantage point for quantitative genetic modelling of complex traits.

From a static to a dynamic view

A static view. Classically, quantitative genetics has assumed a picture of populations that does not include epigenetic variation (FIG. 2a). As a result, traditional models relate phenotypic variation to DNA sequence variation only. Such models form the basis of current QTL techniques (for example, linkage and association mapping), which are geared towards the identification of stable DNA sequence variants that contribute to phenotypes (phQTL^{dna}). Alternative quantitative genetic models could be formulated to relate phenotypic variation to epigenetic variation exclusively (FIG. 2b) — this idea is not too unrealistic, as experimental populations have been constructed that are

Glossary

Chromatin
The nucleoprotein structure that packages DNA within the nucleus of eukaryotic cells. The basic unit of chromatin is the nucleosome: a protein core made up of two molecules each of histones H2A, H2B, H3 and H4, around which 146 bp of DNA is wrapped. Different chromatin states are defined by a range of post-translational modifications of core histones, by incorporation of various histone isoforms as well as by DNA methylation.

Complex traits
Continuously distributed phenotypes that are classically believed to result from the independent action of many genes, environmental factors and gene-by-environment interactions.

Epialleles
Alternative chromatin states at a given locus, defined with respect to individuals in the population for a given time point and tissue type. Epialleles vary greatly in their stability and they affect gene expression levels or patterns rather than gene products.

Epigenetic
Refers to the mitotic or meiotic transmissibility of chromatin variation, independent of DNA sequence variation.

Epigenome
The chromatin states that are found along the genome, defined for a given time point and cell type. Thus, for a given genome there may be hundreds or thousands of epigenomes, depending on the stability of chromatin states.

Epigenotype
The epiallelic constitution of a locus.

epiQTL^{dna}
Refers to a QTL influencing chromatin states (epi) in either *cis* or *trans*, which can be demonstrated to be due to DNA sequence (*dna*).

Genetical genomics
The process of relating DNA sequence variation to molecular profile and phenotypic variation.

Heritability
A concept used in quantitative genetics to denote the proportion of total phenotypic variation in a population that is attributable to variation in the heritable material shared between related individuals.

Nucleolus organizer region
(NOR). A chromosomal region characterized by tandem repeats of ribosomal DNA around which the nucleolus forms.

phQTL^{dna}
Refers to a QTL influencing a phenotype (ph), which can be demonstrated to be due to DNA sequence (*dna*).

phQTL^{epi}
Refers to a QTL influencing a phenotype (ph), which can be demonstrated to be due to chromatin (*epi*).

Tiling array
A subtype of microarray containing small probes that are designed to cover the entire genome or contigs of the genome in an unbiased manner. These arrays can be used coupled with chromatin immunoprecipitation (ChIP–chip), with methyl-DNA immunoprecipitation (MeDIP–chip) and in DNase chip studies.

isogenic (that is, they have almost identical DNA sequences) but nonetheless segregate epigenetic variants[5] (F.J. and V.C., unpublished observations). In these populations, chromatin states can be treated as molecular markers in a genome-wide search for epiallelic determinants of phenotypic variation (phQTL^{epi}). Unlike phQTL^{dna}, which can alter gene products as well as gene expression, phQTL^{epi} are expected to affect mainly gene expression levels or patterns. Clearly, neither of these two extremes — DNA sequence or epigenetic variation alone — is usually encountered in more realistic applications, in which the mapping population is derived from diverse natural or experimental lines. In this situation, the relationship between DNA and chromatin variation can be complicated, not least by the fact that chromatin states can change rapidly within or across generations.

A dynamic view. To address this complication, Richards[7] proposed to conceptualize chromatin-level variation (in this case, at a single locus) in terms of its degree of stability across mitotic and meiotic transitions as well as its level of dependency on DNA sequence variation at the locus or elsewhere in the genome. Based on observations of isolated empirical examples in plants and mammals, he proposed several useful categories of relationship between the genotype and the epigenotype: obligatory, facilitated and pure (FIG. 3a). An obligatory relationship consists of a deterministic association between genotype and epigenotype. Under this arrangement, epialleles are inherited in a stable and strictly sequence-dependent manner across meiosis and mitosis. The influencing DNA sequence variation can act either in *cis* or in *trans*. This obligate link is relaxed in the facilitated category, in which a particular genotype induces changes in epiallelic states in a probabilistic manner, which can then be passed on to subsequent generations. Finally, pure epigenetic variation (which can be further classified as stable or metastable) requires that epigenotypes are completely independent from genotypes.

In light of this classification, the term 'epigenetic', by definition, requires sequence-independent transmission of epialleles, and therefore is only a subset of a variety of other, more dynamic relationships between DNA and chromatin (FIG. 3a) that may harbour important phenotypic consequences. From the perspective of traditional QTL analysis, which relies on sequence-based linkage and association mapping techniques, the most

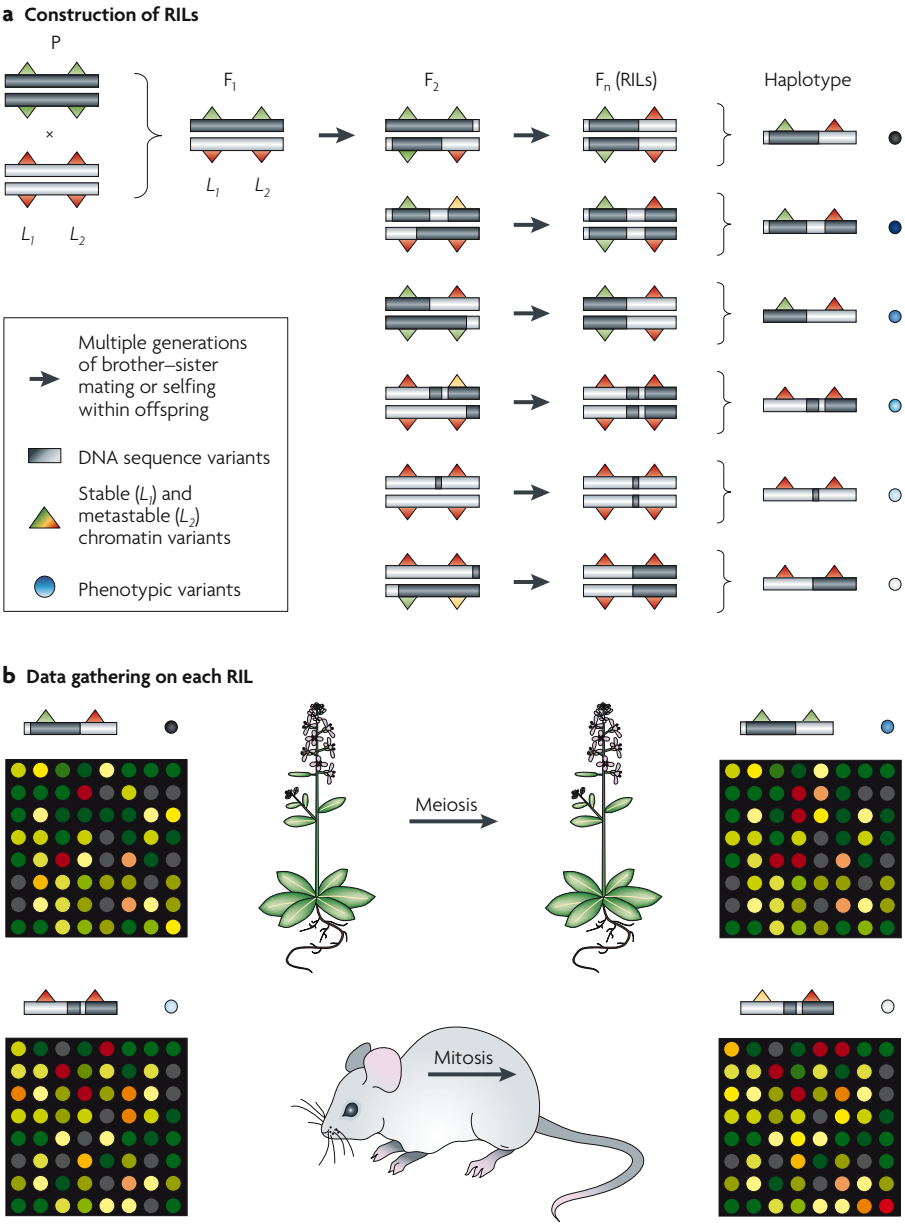

Figure 1 | **DNA sequence and chromatin variation in a segregating population. a** | Two diploid parents (P) with different DNA sequence and chromatin profiles are crossed to generate the filial 1 (F_1) population. Brother–sister mating or selfing generates the filial 2 (F_2) population. Six possible F_2 offspring are shown. Each F_2 individual is the seed of an inbreeding process for multiple generations to generate recombinant inbred lines (RILs). Such lines become fully homozygous after many generations, so that it suffices to show only one haplotype. Of the two epigenomic loci (epi-loci) shown, one (L_1) follows Mendelian inheritance rules, whereas the other (L_2) violates Mendelian inheritance because of metastability (shown as a transition from green (in P) to yellow (in F_2) to red (in RILs)). Horizontal bars indicate the genome, with light and dark grey indicating DNA sequence variants. Triangles indicate chromatin states, with green, yellow and red variants (corresponding to high, intermediate and low gene expression, respectively). For simplicity, only two epi-loci and three levels of chromatin variation are shown. Circles indicate possible phenotypic values of the RILs; RILs are ordered from low phenotypic values (white) to high (dark blue). **b** | Each RIL will be profiled genome-wide using tiling array-based technology for chromatin and phenotypic changes in a time-course experiment that involves either meiotic (top panel) or mitotic (bottom panel) propagation. In both cases, the arrays display instances of both stable and metastable epialleles, whereas the nucleotide sequence in each line remains identical. In a set of RILs, one can therefore study the consequences of epiallelic changes over generational time (top panel) or developmental time (bottom panel) in a range of different DNA backgrounds. The experiment might or might not include environmental intervention to invoke stronger chromatin dynamics.

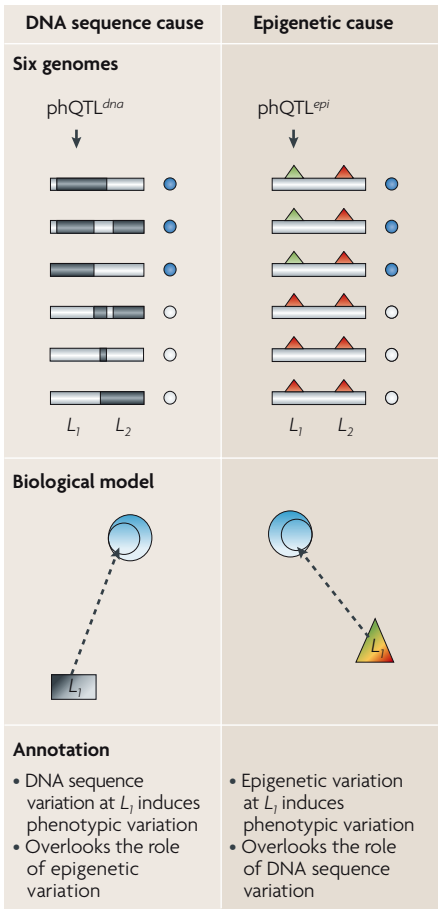

Figure 2 | **Two extreme views of the heritable basis of phenotypic variation.** Two hypothetical populations, one showing DNA sequence variation only (left) and one showing epigenetic variation only (right). The sequence-based QTL analysis (left) detects a QTL (phQTL^{dna}) at locus L_1: the upper three individuals carrying the dark grey DNA sequence variant have higher phenotypic values (blue circles) than the lower three individuals carrying the light grey DNA variant (white circles). The chromatin-based QTL analysis (right) detects a QTL (phQTL^{epi}) at locus L_1: the upper three individuals carrying the green epigenetic variant have higher phenotypic values (blue circles) than the lower three individuals carrying the red variant (white circles). Each of the two separate analyses generates correct but incomplete hypotheses about the causal architecture underlying phenotypic variation in the more realistic cases of a single population in which these two sources of variation co-occur. A full analysis of this example is given in FIG. 4. The biological model shows hypothesized connections between phenotypic variation and its heritable basis. Blue circles indicate phenotypic variation: the inner blue circle indicates the proportion of variation explained by DNA sequence (left) or epigenetic (right) variation, whereas the outer blue circle indicates the total variation, including the influence of other factors on phenotype.

problematic situation consists of the partial or complete uncoupling of DNA and chromatin variation (FIG. 3a; facilitated or pure epialleles). This is because phenotypes are related to epigenetic variation but are only partly related to DNA sequence variation, and sequence-based QTL estimates do not capture this source of heritable variation. This shortcoming might be reflected in the frequent observation that heritability calculations, which make no assumptions about the molecular determinants of phenotypes, exceed the sum of estimated QTL effects; although other issues, such as low statistical power to detect phQTL^{dna} with small effect sizes and epistasis, certainly also have a role. Another challenging situation relates to chromatin variation that is completely determined (in *cis* or in *trans*) by DNA sequence variation (FIG. 3a; obligate epialleles). This is problematic as it raises the important question of how many of the QTLs detected by sequence-based methods, and for which variation does not affect genes encoding products involved in chromatin control (for example, DNA or histone methyltransferases), are nonetheless attributable to chromatin effects[30].

Experimental analysis and implementation

Mapping* cis- *and* trans-*acting factors. In the analysis of real data, the above classification will probably not be encountered as discrete categories, but rather as particular instances of a continuum of statistical relationships between sequence alleles and epialleles. With genome-wide sequence and chromatin data obtained from each individual and a suitable probability model we can begin to formally classify epialleles along the epigenome according to their degree of dependence on DNA variation in *cis* or in *trans*, and on their level of stability as exemplified by changes in QTLs and epiallelic covariance information (FIG. 3b,c). Such an analysis will provide a comprehensive inventory of the likely prevalence of different types of epialleles as well as their genomic distribution. The result can be interpreted as an epigenomic reference map for a particular population under consideration, which is annotated according to its dependence on sequence variation. An advantageous feature of using a QTL approach in this setting is the possibility of identifying novel loci involved in chromatin control. In a previous study of local DNA methylation, none of the known genes involved in *de novo* establishment and maintenance of DNA methylation mapped to the region of one of the identified QTL intervals[3]. Hence, further fine-mapping and sequencing of the QTL interval will eventually yield the causal DNA variants. The results from the proposed epigenome QTL analysis can be shown in so-called '*cis–trans* plots' (FIG. 3c). We highlight separately those categories of epigenomic loci (epi-loci) that would show a heritable effect on the phenotype (FIG. 3d). The classification outlined in FIG. 3 summarizes concisely the types of decomposition that can be achieved.

Design considerations. Although the conceptual thrust of the proposed approach is novel, its implementation is largely consistent with current genetical genomics studies, which aim to relate DNA sequence variants to genome-wide expression profiles. In recent years these studies have been successfully used in both experimental and natural populations[31]. Specific methodological solutions that have been developed for these studies[32], for example, power calculations and array hybridization designs, can therefore be easily extended and used in the planning of the proposed experiments, in which gene expression data will be replaced with array-based chromatin data. This convenient feature should prevent design issues from becoming major obstacles when executing the proposed approach.

Several authors have gone one step further and have discussed the inclusion of various environmental perturbation regimes in genetical genomics studies[33–35]. These considerations might prove particularly important in the present context to unravel the impact of external factors, for example, cold temperature treatments, dietary changes and radiation exposure, on the dynamic link between genotype, epigenotype and phenotype over generational or developmental time[36].

Promising applications

Meiotic example applications. The relevance of epigenetic variation in studies of heritable phenotypes is probably organism dependent. In plants, epialleles, such as those associated with differences in DNA methylation, can be remarkably stable and are more readily carried over to subsequent generations. In mammals, however, epialleles are believed to be erased during gametogenesis or early development. Apart from furthering our basic understanding of epigenetic inheritance, the genome-wide isolation of sequence-independent, stable epialleles could be an important goal in marker-assisted plant breeding programmes. The incorporation of both DNA sequence and chromatin information might

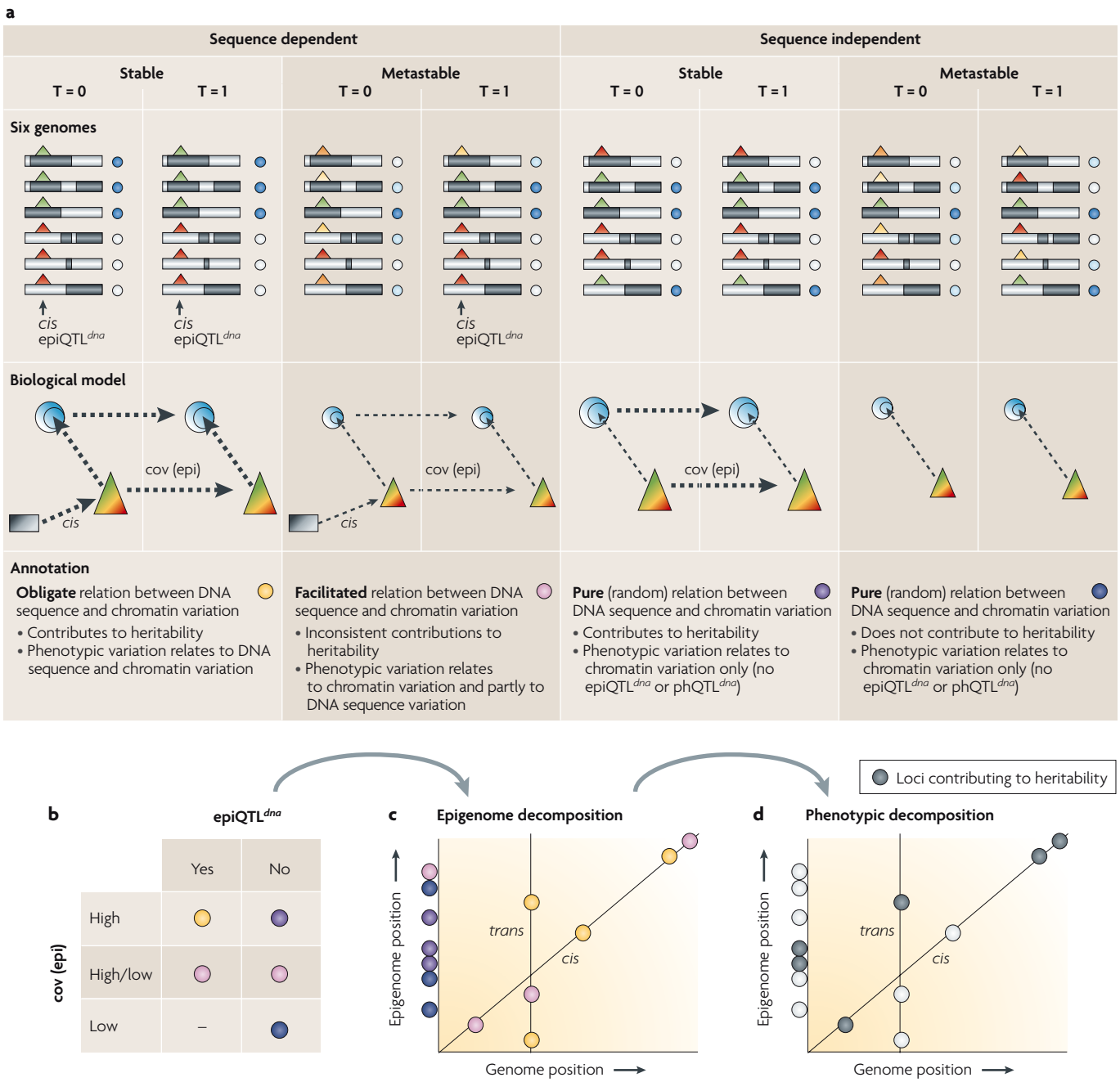

Figure 3 | **Classification of epialleles. a** | Six recombinant inbred lines (RILs) are shown at two meiotic or mitotic generations (T = 0 and T = 1) for each of the four classes of epialleles at a single epigenomic locus (epi-locus). See the main text for a description of the classification of epialleles into obligate, facilitated and pure. Horizontal bars indicate the genome, with light and dark grey areas indicating DNA sequence variants. Triangles indicate chromatin states, with green, orange/yellow and red variants corresponding, respectively, to high, medium and low gene expression; circles indicate low (white) and high (dark blue) phenotypic values. The biological models illustrate the hypothesized relationships from QTL analysis between DNA sequence, chromatin and phenotypic variation. Thick dotted lines indicate stronger covariance (cov). Epialleles contribute to different extents to phenotypic heritability in the first three situations shown, but not at all in the last one (right), because of extreme instability (indicated by transitions from red to yellow, orange to green and yellow to red). **b** | The chromatin data obtained at different time points can be correlated. With a suitable probability model (for example, a multiple trait model) this covariation information can be combined with the detection (or lack thereof) of an epiQTL^{dna} to classify the epialleles. For example, a 'high' covariation in chromatin states at a given locus provides evidence that epialleles have segregated in a stable manner (yellow circles). Moreover, if we also detect an epiQTL^{dna} that influences the epiallelic states at the locus (that is, its covariation structure) in *cis* or in *trans*, we can further conclude that the stable transmission of these epialleles is sequence dependent (yellow circles). **c** | The results of a QTL analysis can be visualized in a so-called *cis–trans* plot. In this plot, all the data points that fall inside of the plot correspond to epialleles that are transmitted in a sequence-dependent manner. That is, their transition is 'constrained' by the genotypic context locally (*cis*) or distally (*trans*). On the other hand, the data points that fall outside of the plot (along the y-axis) represent epialleles that are sequence-independent, and are either transmitted or not transmitted. **d** | This graph shows the same as in panel **c**, but those categories of epi-loci that would show a heritable effect on the phenotype (grey circles) are highlighted separately.

be a promising route towards commercially superior phenotypic outcomes. The quantitative genetic approach outlined here for RILs applies to many types of populations used in breeding.

However, even in mammals, single-locus examples indicate that epiallelic states can escape erasure and can be subject to transgenerational inheritance[6,7]. To what extent this occurs on a genome-wide scale remains unknown. In multifactorial human diseases such as diabetes or heart disease, in which each gene makes small contributions to an underlying continuous predisposition, the fidelity of epiallelic transmission cannot be directly inferred from observations of phenotypes, as might be the case with more extreme, qualitative traits. It therefore seems necessary to assess the transmission of epialleles in a locus-by-locus manner using genome-wide analysis over multiple generations to fully grasp the heritable architecture of these complex diseases. Such studies can be done with animal models, for example, using existing mouse or rat RILs, or even outbred mammalian populations, for example, humans or livestock.

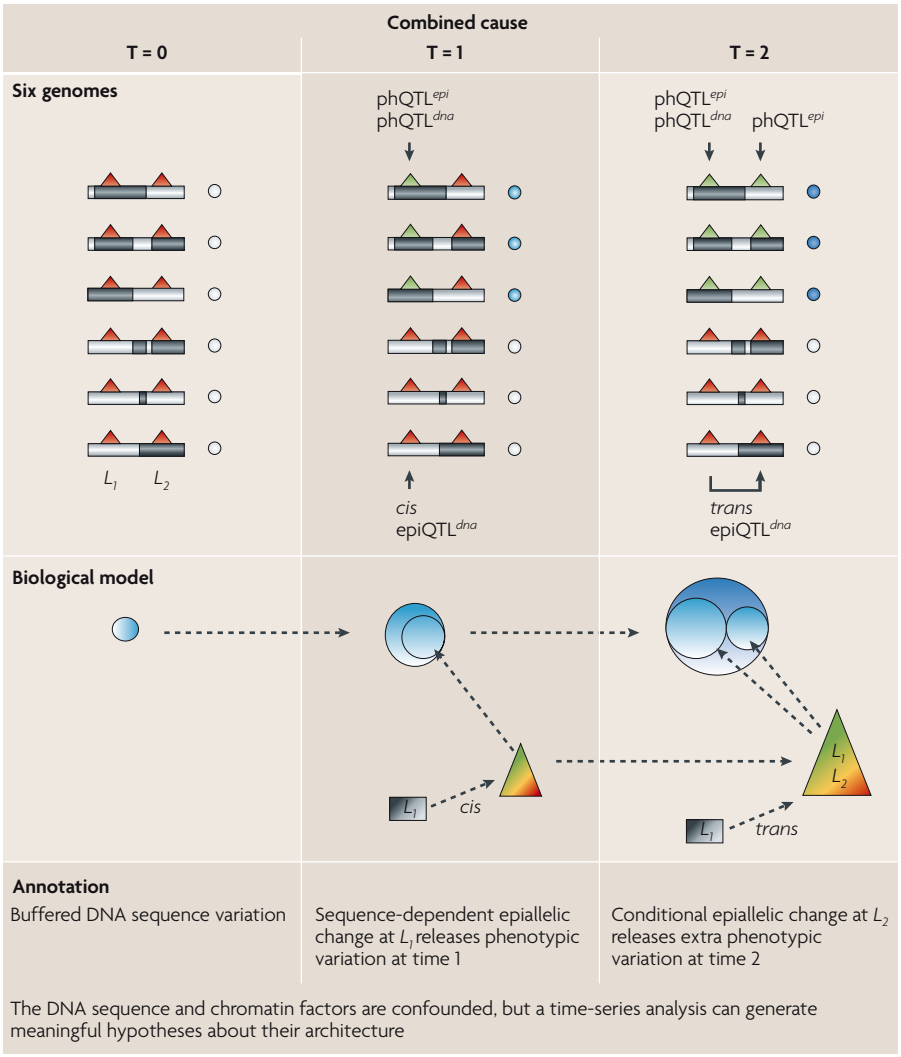

Figure 4 | **Phenotypic variation: a complex case.** Although separate analyses using either DNA sequence or chromatin information in QTL mapping can generate incomplete hypotheses (as shown in FIG. 2), QTL analysis can generate meaningful hypotheses about the causal architecture underlying phenotypic variation. Horizontal bars indicate the genome, with light and dark grey DNA sequence variants. Triangles indicate chromatin states, with green and red variants corresponding, respectively, to high and low gene expression; circles indicate low (white) and high (dark blue) phenotypic values. At the first time point (T = 0), a QTL scan based on DNA or chromatin markers will not result in the detection of any QTL at any of the two loci (L_1 or L_2). This is because the presence of non-polymorphic epigenomic loci (epi-loci) suppresses or 'buffers' differential gene expression, thereby preventing DNA sequence variation from becoming phenotypically manifest. Assume an environmental perturbation at T = 0 that affects L_1, so that in the next generation (when T = 1) a sequence-dependent (*cis*) chromatin change takes place, which releases buffered sequence variation. At T = 1, a QTL search based on DNA markers will now yield a phQTL^{dna} but will also give a phQTL^{epi} if the search is extended to chromatin markers. Note the unexpected gain in phenotypic variation in the population at this generation through the release of previously hidden sequence variants. This effect is further attenuated if we proceed to the generation at T = 2. At this stage, locus L_1 induces chromatin changes in *trans* at locus L_2. The biological model visualization uses symbols as defined in FIGS 1–3. The increased size of the blue phenotypic circles illustrates the release of phenotypic variation over time as the result of *cis* and *trans* effects of DNA variation on chromatin variation. Arrows indicate the relationships inferred from QTL analyses.

Mitotic example applications. The dynamic aspect of chromatin variation has received far more attention in developmental contexts, that is, across mitotic transitions during the lifetime of the organism. Mammalian cancer research, in particular, has demonstrated an interest in processes such as accidental loss or gain of DNA methylation — so-called epimutations — which has become a useful biomarker for aberrant cellular development, that is, cancer[37]. Similar but more orchestrated chromatin changes on the level of DNA and histone methylation also have an important role in cell lineage determination during normal development in the mouse[38–40]. Moreover, a cellular comparison between two different inbred mouse strains revealed notable allele specificity in histone methylation, suggesting obligatory or facilitating *cis* influences of DNA sequence variation in this case[40].

Both types of application could greatly benefit from a more integrative analysis using advanced experimental or natural crosses, or cell lines of these crosses, to assess the impact of sequence variation, particularly in *trans*, on developmentally driven chromatin changes. Arguably, genotype-by-epigenotype interactions could be of particular importance there[41]. The proposed decomposition can be used as a global screening tool to identify sequence contexts (*cis* or *trans*) in which epiallelic transitions occur more readily.

Buffering and release of DNA sequence variation. The value of the proposed approach is also reflected in more complex intragenerational or transgenerational applications in which changes in the epiallelic structure in a population can give rise to the buffering or release of pre-existing DNA sequence variation (FIG. 4). Such phenomena can occur

when chromatin changes interact with DNA variants through the silencing or activation of genes. Intriguing, although indirect, evidence for time-dependent genetic effects has been reported in several developmental QTL studies[42–45]. For instance, in an early experiment Cheverud[45] examined body weight in mice over the course of 10 weeks, and discovered different sets of QTLs for the early and later growth stages. The authors concluded that these results are consistent with different genetic and physiological systems being causally active in an age-dependent manner. Comparable results have been obtained in a genome-wide scan for imprinted QTLs, suggesting that these types of parent-of-origin QTLs are under similar developmental control[46].

It remains an open empirical question whether analogous processes can operate transgenerationally. A particularly controversial idea suggests that environmental perturbations can evoke meiotically transmissible chromatin changes[7,36] which, in turn, uncover previously hidden DNA variants in the population. This is an important consideration for ecology and evolutionary biology as it describes a mechanism by which phenotypically relevant sequence variation is 'created' without requiring any novel mutational input.

A comprehensive analysis of the relationship between genotype, epigenotype and phenotype, coupled with systematic environmental perturbation regimes, can provide a means to explore these questions and to generate meaningful hypotheses about the physical basis of such intragenerational or transgenerational phenomena (FIG. 4).

Discussion

Current sequence-based QTL approaches for dissecting complex traits into its heritable components do not consider epiallelic variation. This neglect can have far-reaching implications, as these studies might miss important heritable phenotypic effects exerted by epigenetic variants (FIG. 2). Moreover, chromatin changes induced by environmental or genomic perturbations can lead to short-term or longer-term interactions with existing DNA sequence variation in populations through buffering and release processes[47]. These considerations point towards a heritable architecture that is both more complex and more dynamic than previously appreciated. If this can be empirically verified, we might be forced to re-evaluate our current models of the mode and tempo of adaptive processes in natural populations, as was attempted in several early theoretical studies[48,49].

Ultimately, we will be confronted with the difficult task of defining the properties of epialleles in populations. It is likely that such a definition will need to be contextualized in terms of conditional dependencies on environmental and DNA sequence variables. In this case, it will be challenging to find a suitable level of abstraction that allows for a meaningful exploration of the merger between classical sequence-based quantitative genetics and epigenome dynamics. We argue that the most crucial considerations will probably relate to the function of time that governs epiallelic transitions both within and across generations[50], the cell or tissue types used for measurements, the proper functional units of analysis for defining an epiallele (for example, a single cytosine versus composite measurements of DNA methylation over promoter sequences), as well as the specific contextual properties of genomic regions (for example, heterochromatin versus euchromatin) and of the environment (for example, stressful versus non-stressful). The experimental strategy proposed here (FIG. 1) can serve as a starting point to explore some of these issues in empirical data. Supplemental pedigree designs (for example, reciprocal crosses) might eventually be required to effectively distinguish detected epialleles from parentally imprinted alleles, particularly in mitotic experiments. This point needs careful consideration in mammals, in which imprinting patterns are established in the germ line of the parents and maintained in somatic cell lineages of the progeny throughout development and adult life[46]. The conceptual and experimental framework presented in this article should advance our basic understanding of complex traits, and should therefore be of broad appeal to a range of fields, including agriculture, biomedicine and evolution.

In summary, the integration of chromatin-level data in quantitative genetic studies poses formidable challenges at the forefront of multidisciplinary research, but promises to significantly alter our view of how phenotypes are created and sustained in populations over time.

Frank Johannes and Ritsert C. Jansen are at the Groningen Bioinformatics Centre, University of Groningen, NL-9751 NN, Haren, The Netherlands.

Vincent Colot is at the National Centre for Scientific Research UMR8186, Department of Biology, Ecole Normale Supérieure, 75230 Paris Cedex 05, France.

Ritsert C. Jansen is also at the University Medical Centre Groningen, Department of Genetics NL-9700 RB, Groningen, The Netherlands.

Correspondence to F.J. e-mail: f.johannes@rug.nl

doi:10.1038/nrg2467

1. Vaughn, M. W. *et al.* Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.* **5**, e174 (2007).
2. Zhang, X., Shiu, S., Cal, A. & Borevitz, J. O. Global analysis of genetic, epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana* using whole genome tiling arrays. *PLoS Genet.* **4**, e1000032 (2008).
3. Riddle, N. C. & Richards, E. J. The control of natural variation in cytosine methylation in *Arabidopsis*. *Genetics* **162**, 355–363 (2002).
4. Riddle, N. C. & Richards, E. J. Genetic variation in epigenetic inheritance of ribosomal RNA gene methylation in *Arabidopsis*. *Plant J.* **41**, 524–532 (2005).
5. Kakutani, T., Munakata, K., Richards, E. J. & Hirochika, H. Meiotically and mitotically stable inheritance of DNA hypomethylation induced by *ddm1* mutation of *Arabidopsis thaliana*. *Genetics* **151**, 831–838 (1999).
6. Peaston, A. E. & Whitelaw, E. Epigenetics and phenotypic variation in mammals. *Mamm. Genome* **17**, 365–374 (2006).
7. Richards, E. J. Inherited epigenetic variation — revisiting soft inheritance. *Nature Rev. Genet.* **7**, 395–401 (2006).
8. Henderson, I. R. & Jacobsen, S. E. Epigenetic inheritance in plants. *Nature* **447**, 418–424 (2007).
9. Chan, S. W., Henderson, I. R. & Jacobsen, S. E. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Rev. Genet.* **6**, 351–360 (2005).
10. Martienssen, R. A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070–1074 (2001).
11. Stam, M., Belele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* **16**, 1906–1918 (2002).
12. Soppe, W. J. *et al.* The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791–802 (2000).
13. Das, O. P. & Messing, J. Variegated phenotype and developmental methylation changes of a maize allele originating from epimutation. *Genetics* **136**, 1121–1141 (1994).
14. Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I. & Whitelaw, E. Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002).
15. Penterman, J. *et al.* DNA demethylation in the *Arabidopsis* genome. *Proc. Natl Acad. Sci. USA* **104**, 6752–6757 (2007).
16. Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C. & Paszkowski, J. Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* **130**, 851–862 (2007).
17. Richards, E. J. Population epigenetics. *Curr. Opin. Genet. Dev.* **18**, 221–226 (2008).
18. Zilberman, D. & Henikoff, S. Genome-wide analysis of DNA methylation patterns. *Development* **134**, 3959–3965 (2007).
19. Schones, D. E. & Zhao, K. Genome-wide approaches to studying chromatin modifications. *Nature Rev. Genet.* **9**, 179–191 (2008).
20. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Rev. Genet.* **9**, 465–476 (2008).
21. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
22. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37**, 853–862 (2005).
23. Zhang, X. *et al.* Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* **5**, e129 (2007).
24. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006).
25. Zilberman, D. The human promoter methylome. *Nature Genet.* **39**, 442–443 (2007).
26. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39**, 61–69 (2007).
27. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **2**, 395–397 (2008).

28. Jansen, R. C. Studying complex biological systems using multifactorial perturbation. *Nature Rev. Genet.* **4**, 145–151 (2003).
29. Mager, J. & Bartolomei, M. S. Strategies for dissecting epigenetic mechanisms in the mouse. *Nature Genet.* **37**, 1194–1200 (2005).
30. Petronis, A. Epigenetics and twins: three variations on the theme. *Trends Genet.* **22**, 347–350 (2006).
31. Rockman M. V. & Kruglyak L. Genetics of global gene expression. *Nature Rev. Genet.* **7**, 862–872 (2006).
32. Rosa G. J.M, *et al.* Review of microarray experimental design strategies for genetical genomics studies. *Physiol. Genomics* **28**, 15–23 (2006).
33. Li Y. *et al.* Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* **2**, e222 (2006).
34. Li Y. *et al.* Generalizing genetical genomics: getting added value from environmental perturbation. *Trends Genet.* **24**, 518–524 (2008).
35. Smith E. N. & Kruglyak L. Gene–environment interaction in yeast gene expression. *PLoS Biol.* **6**, e83 (2008).
36. Bossdorf, O., Richards, C. L. & Pigliucci, M. Epigenetics for ecologists. *Ecol. Lett.* **11**, 106–115 (2008).
37. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).
38. Meissner *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
39. Farthing *et al.* Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **4**, e1000116 (2008).
40. Mikkelsen *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
41. Bjornsson, H. T., Fallin, M. D. & Feinberg, A. P. An integrated epigenetic and genetic approach to common human disease. *Trends Genet.* **20**, 350–358 (2004).
42. Johannes, F. Mapping temporally varying quantitative trait loci in time-to-failure experiments. *Genetics* **175**, 855–865 (2007).
43. Wu, R. & Lin, M. Functional mapping — how to map and study the genetic architecture of dynamic complex traits. *Nature Rev. Genet.* **7**, 229–237 (2006).
44. Henckaerts, E., *et al.* Genetically determined variation in the number of phenotypically defined hematopoietic progenitor and stem cells and their response to early acting cytokines. *Blood* **99**, 3947–3954 (2002).
45. Cheverud, J. M., *et al.* Quantitative trait loci for murine growth. *Genetics* **142**, 1305–1319 (1996).
46. Wolf J. B., *et al.* Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLoS Genet.* **4**, e1000091 (2008).
47. Sollars, V. *et al.* Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nature Genet.* **33**, 70–74 (2003).
48. Csaba, P. Plasticity, memory and the adaptive landscape of the genotype. *Proc. R. Soc. Lond. B* **265**, 1319–1323 (1998).
49. Pal, C. & Miklos, I. Epigenetic inheritance, genetic assimilation and speciation. *J. Theor. Biol.* **200**, 19–37 (1999).
50. Rando, O. J. & Verstrepen, K. J. Timescales of genetic and epigenetic inheritance. *Cell* **128**, 655–668 (2007).

Acknowledgements

We would like to thank R. Breitling for critical comments on previous versions of the manuscript, and Y. Li for her help in preparing the figures. The F.J. and R.C.J. group is supported by the Netherlands Organization for Scientific Research (NWO-ALW VICI grant). V.C. is a NET member of the European Union Epigenome Network of Excellence and is supported by the National Centre for Scientific Research (CNRS), Génoplante and the French Agence Nationale de la Recherche (ANR).

FURTHER INFORMATION

Arabidopsis Epigenetics and Epigenomics, CNRS: http://www.biologie.ens.fr/a2e
Epigenome Network of Excellence: http://www.epigenome-noe.net
Groningen Bioinformatics Centre: www.rug.nl/gbic

ALL LINKS ARE ACTIVE IN THE ONLINE PDF